|  |  |  |  |
| --- | --- | --- | --- |
| Title | HESIP: A Hybrid System for Explaining Sub-symbolic Predictions | | |
| Author(s) Name | Abdus Salam, Rolf Schwitter, Mehmet A. Orgun | | |
| Contact Email(s) | abdus.salam@aiub.edu | | |
| Published Journal Name | The 34th Australasian Joint Conference on Artificial Intelligence | | |
| Type of Publication | Conference Article | | |
| Volume | 13151 | Issue |  |
| Publisher | Springer, Cham | | |
| Publication Date | 19 March 2022 | | |
| ISSN |  | | |
| DOI | https://doi.org/10.1007/978-3-030-97546-3\_3 | | |
| URL |  | | |
| Other Related Info. | Pages 27–39 | | |
|  | | | |

|  |  |
| --- | --- |
| Abstract |  |
| Machine learning models such as neural networks have been successfully used in many application domains such as mission critical systems, digital health and autonomous vehicles. It is important to understand why particular predictions are made by a sub-symbolic machine learning (ML) model, because humans use these predictions in their decision making process. In this paper, we introduce HESIP, a hybrid system that combines symbolic and sub-symbolic representations to explain a prediction in natural language for an image prediction task. A sub-symbolic ML model makes a prediction for an image, and based on this predicted image, the system selects sample images from the dataset. Afterwards, a symbolic ML model learns probabilistic rules using the representation of positive and negative sample image instances where the decision about a positive or negative image instance comes from the sub-symbolic ML model. The prediction of an image can then be explained in natural language using the learned rules. Our evaluation shows that the probabilistic rules can be learned with high accuracy. | |